Track & Know | Issue 8 | Date: January 2021

WELCOME TO THE FINAL TRACK & KNOW NEWSLETTER!

In this eigth newsletter, you can find:

- A look back at the Track & Know final event
- What we have achieved: a project round-up
- Project brochure available
- Discover the Track & Know YouTube channel

ABOUT THE NEWSLETTER

This newsletter informs you about the results and activities of the EU H2020 research project Track& Know. The aim is to keep all relevant actors interested in managing big data, more specifically in the type of big data we focused on in the project and the tools/methods we developed to handle, analyse and visualize these datasets. T&K focused on resolving key business cases for 3 test pilots, namely transport/mobility, insurance and health care. Business cases which were explored in these pilots are as follows but not limited to: minimizing patients travel, carpooling and electric mobility potential, driver behaviour profiling etc.

Project coordinator
Dr. Ibad Kureshi
Senior Research Scientist – Inlecom Systems
Square de Meeus 38/40
1000 Brussels
Belgium
Ibad.kureshi@inlecomsystems.com

Track & Know final event @ European Big Data Value Forum 2020

Early November, Track & Know's final event took place during the **European Big Data Value Forum 2020 conference**. Track & Know was involved in **4 sessions** during this 3-day flagship event of the European Big Data and Data-Driven AI Research and Innovation community organized by the Big Data Value Association (BDVA) and the European Commission (DG CNECT).

EUROPEAN BIG DATA VALUE FORUM
3 - 5 NOVEMBER . 2020 - BERLIN + VIRTUAL

With a central theme "**Building a strong European Data and AI Ecosystem**", the EBDVF 2020 edition brought together European communities on AI and Data. The event also focused on how Data and AI can help to tackle emerging societal challenges, like those brought by the Covid19 pandemic.

Would you like to take another look at our 4 sessions? **The recorded sessions are available online.**

1. Parallel session on European Big Data Research for industry. 3 projects, 7 sectors, 9 applications, 41 software components. Now what?

Watch this session

The Big Data research projects **Track and Know** (Big Data for Mobility Tracking Knowledge Extraction in Urban Areas), **BigDataStack** (High performance data-centric stack for Big Data applications and operations) and **I-BiDaaS** (Industrial-Driven Big Data as a Self-Service Solution) hosted a joint session bringing together their findings in terms of barriers to adoption of Big Data research in different sectors as well as in terms of current and future impact of their research.

Speakers also elaborated on the concrete business questions that have been answered in the project pilots. For Track & Know, **Toni Staykova** from UKeMED was a member of the expert panel.

2. Sponsor talk Track & Know – Big Mobility Data Integration platform – approach, development and application.

Watch this session

This presentation by **Marios Logothetis** from Intrasoft International aimed at presenting the Track and Know project platform and at sharing and showcasing a robust and scalable platform for mobility data. The presentation included screen shots/demo of the T&K Big Data platform.

3. Evaluation schemes for Big Data and AI Performance of high business impact.

Watch this session

The first part of this session focussed on **Big Data Benchmarking landscape and Big Data Pipelines. Athanasios Koumparos** from Vodafone Innovus presented the Track & Know project, its challenges and results. This session was also the closing event of liaison project DataBench.

EUROPEAN BIG DATA VALUE FORUM
3 - 5 NOVEMBER . 2020 - BERLIN + VIRTUAL

4. Application Track 3 – Transport, Mobility and Logistics – Main session

Watch this session

Download the presentation

The transformation of the multimodal transportation ecosystem is happening now – changes are rapid and involve all modes of transportation. New players use the power of real-time data to offer **personalized door-to-door travel and logistics services**. At the same time there is an expressed **need for an economical, flexible, safe, and more sustainable way to get from A to B by road, rail or water – using all available innovative ways** since:

- in most EU countries entrants are challenging the existing transport practices and services naming them as unable to fully satisfy a never-ending demanding audience;
- providers leverage the use of mobiles to create new relationships with travelers;
- societies have to reduce their CO2 footprint,
  thus, sustainability becomes an important additional aspect of mobility and our mobility concepts;

Big Data, Data Issues and Artificial Intelligence can play an important part in achieving this vision. **Yannis Theodoridis** from UPRC joined this session with his presentation which is entitled 'Learning from our movements'. He elaborated on the T&K objectives, datasets and toolboxes.

Key achievements of the Track & Know project

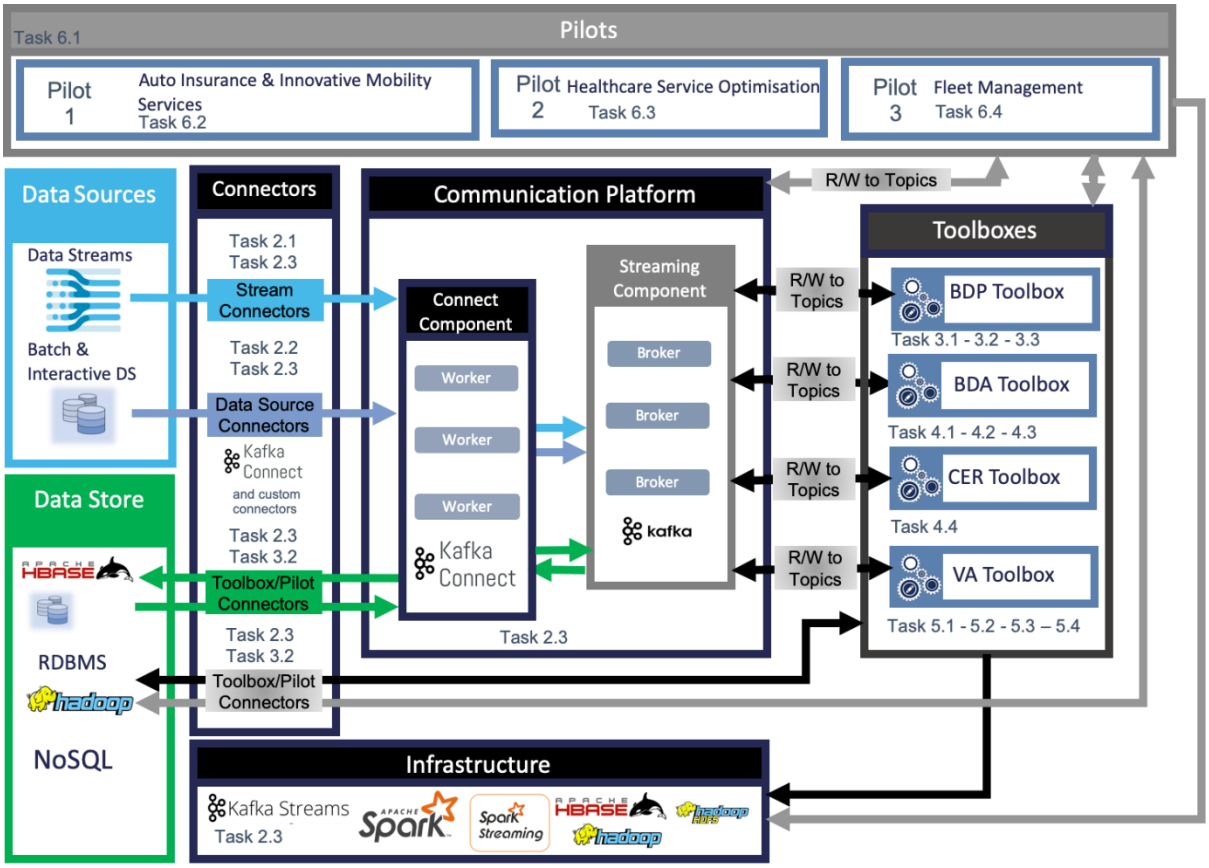
The project primarily focused on mobility-based big data and developed innovative software stacks and toolboxes that addressed key questions in emerging cross-sector markets such as commercial IoT services, car insurance and healthcare management.

It started with a background that existing and constantly generated new mobility data (big data) largely remains decoupled and isolated within private infrastructures. **Efficient handling and inferring knowledge by effectively blending such big data can open new opportunities for business in particular and societies in general.** In today's society, the ability to retrieve knowledge from mobility and contextual data is becoming more and more critical for the competitiveness of all the economic, political and cultural entities. Multidisciplinary research teams from Mobility Data management, Complex Event Recognition, Geospatial Modelling, Complex Network Analysis, Transportation Engineering and Visual Analytics as part of a Track & Know project consortium developed new big data management platform and various toolboxes for processing and analysing mobility big data. Additionally, to demonstrate the applicability of a developed platform and toolboxes, the project also focused on resolving key business cases for 3 test pilots, namely transport/mobility, insurance and health care. Business cases explored in these pilots are as follows but not limited to: minimizing patients travel, accident risk estimation, carpooling and electric mobility potential, driver behaviour profiling etc.

1. Big Mobility data Integrator (BMDI)

It is a Kafka based, horizontally scalable, distributed streaming platform. It facilitates the efficient, high performance and horizontally scalable ingestion of data from a variety of data sources while ensuring its persistence and availability through a multi-client publish/subscribe model. It enables the decoupling and straightforward integration of various processing and toolbox components whilst ensuring secure communications with load balancing characteristics. It advances parallel processing capabilities by introducing data partitioning and subsequent processing by multiple instances of toolbox code deployed on the required hosts, enabling the setup of Big Data Processing Pipelines.

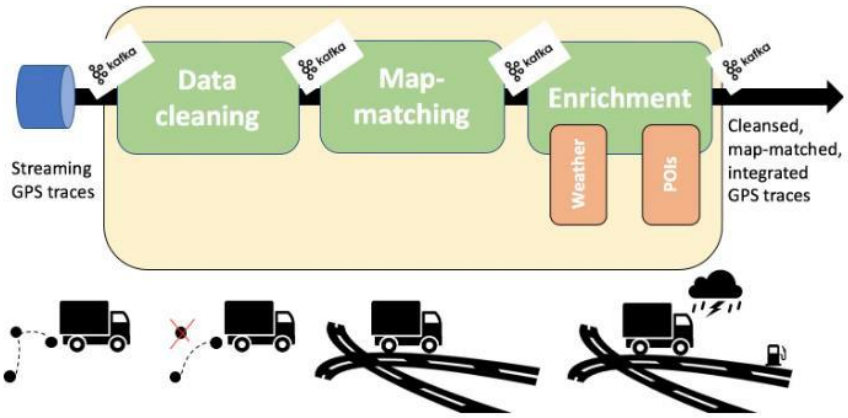
It Introduces Kafka Connect functionality where a variety of Custom and Community available connectors can be utilised. It provides a set of GUI components for Platform administration and necessary dashboards for monitoring various performance-related metrics. The BMDI platform is able to efficiently interoperate with modern data storage technologies of a Big data ecosystem such as RDBMS, NoSQL, HDFS Hadoop, Apache HBASE, etc. as well as other persistence approaches such as Mongo, MySQL, JDBC, etc. Big data applications (toolboxes) can be implemented in all-important Big Data languages including Python, Java, R and Scala. Traditional programming approaches (C/C++, Ruby, Perl, PHP) are also supported. Big data toolboxes developed within the project such as big data processing (BDP) toolbox, big data analytics (BDA) toolbox, Complex event recognition (CER) toolbox and Visual analytics (VA) toolbox are integrated into the platform to demonstrate the business cases of 3 pilots.



2. Data processing and enrichment pipeline

A data pipeline is developed as part of the BDP toolbox as shown in the figure. It processes the raw GPS data and provides in its output cleansed trajectories of moving objects. The pipeline uses processes such as data cleansing, annotation, interpolation, map-matching and enrichment of data with the weather (e.g. wind, rain, ice etc.) and Point-of-interest (e.g. fuel station etc.) information.

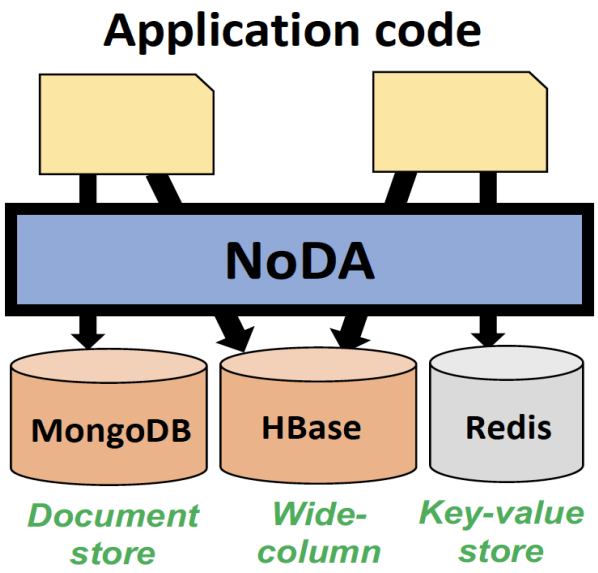
This is developed with a viewpoint to support processes in BDA toolbox, to help answer questions on accident risk estimation, hot spot analysis, electric mobility analysis.



The salient features of the enrichment architecture include operating in an online manner and being reusable across diverse mobility data (urban, maritime, air-traffic). Furthermore, the developed enrichment approach can handle more complex geometries than simple positions (e.g., associating weather with a 3D sector), and produce output in Resource description framework (RDF).

3. Unified NoSQL Data Access Operators—the NoDA API

Track & Know researchers developed an approach within a BDP toolbox which is an abstraction layer in the form of an API over the NoSQL stores. The work carried out contains: (a) the specification of an API for data access/retrieval operators over NoSQL stores, and (b) its implementation for a specific NoSQL store, namely MongoDB, HBase and Redis.

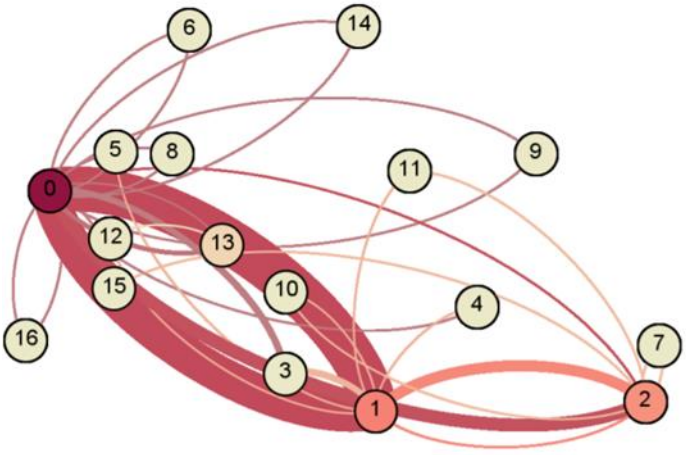


The designed API for big data access operators can be used by big data developers for easy access to NoSQL stores while hiding implementation details relevant to the specific NoSQL store. This allows developers to write programs against this API, which is much easier, more comprehensive, and removes the "entry barrier" of having to learn the internal operation of a specific NoSQL store. NoDA is JVM-compliant and can return results as Spark Data frame.

4. Extraction and Semantic Annotation of Individual Mobility Networks (IMN) for the Purpose of Crash Risk Prediction

IMNs are a concise graph representation of the mobility history of individuals. From raw GPS traces, the trajectories of a single mobility user are reconstructed and processed to infer the relevant locations that the user visited (the nodes of IMNs) and aggregate the trips between two locations (the edges of IMNs). Several features are extracted that will be used as key variables in the model. These are as follows: Number of nodes and edges, Node and Edge Frequency, clustering co-efficient of IMN, Network modularity index, Measure of IMN change with respect to time that reflect changes in habits or mobility needs.

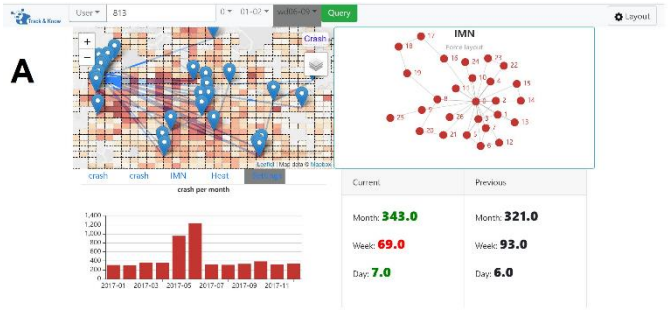
As part of BDA toolbox, a predictive model for crash risk is developed using a machine learning technique. The model uses IMNs along with other features of an individual mobility which are correlated to events that happen in the mid/long-term future, such as crashed in the next month. The component is implemented in Python, is completely automatized and takes as input the mobility data of the user in the monitoring period, as well as (only for the construction of the model, and not for its usage in prediction) a label specifying whether the event is going to happen.



5. Track & Know dashboards

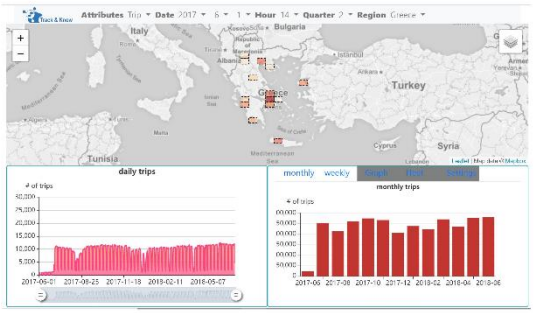
As part of the VA toolbox, T&K dashboard provides an overview of some aggregated statistics, as well as the analytical results generated from the other toolboxes. In short, the dashboard does not serve as an interface of the analytical tools but rather visualizes the analytical results, which is generated from off-line big data analytics (thus, it acts as a “strategic dashboard”). It provides simple interactions (e.g., selection, filtering, and drilling down), allowing users to explore the analytical results on the browser side, partially as an analytical dashboard. It is implemented using web technologies and can be accessed via web browsers to maximize accessibility by end-users. In the background, it connects to the database to receive real-time updates of aggregated statistics and analytical results and visualize them on the dashboard interface. The outputs are planned to use the Kafka platform as the intermediate while the dashboard will consume these outputs from Kafka.

A

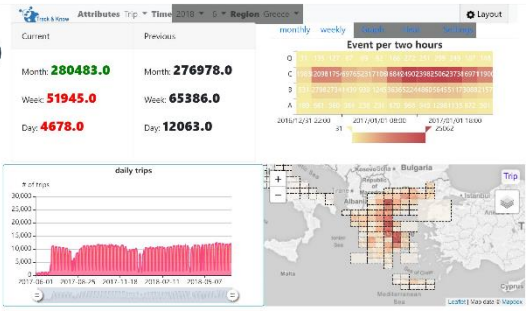


B

C



D



The above methods/tools/functionalities are just a few examples of developments/achievements made in the project. There are demonstrating videos available on where these methods/tools are utilised in addressing a particular research question for the 3 pilots’ cases. Interested readers are requested to visit the Track and know website to explore more developments and demo videos. In addition to this, the website also contains an **online observatory**, where users can request and explore the sample data sets, download the developed software packages and other resources generated within the project.

Track & Know brochure & YouTube channel available

The Track & Know project brochure is available online. This summary publication gathers the **most important results and achievements** of this H2020 project on efficient and effective Big Data handling.

[Discover the project brochure now]

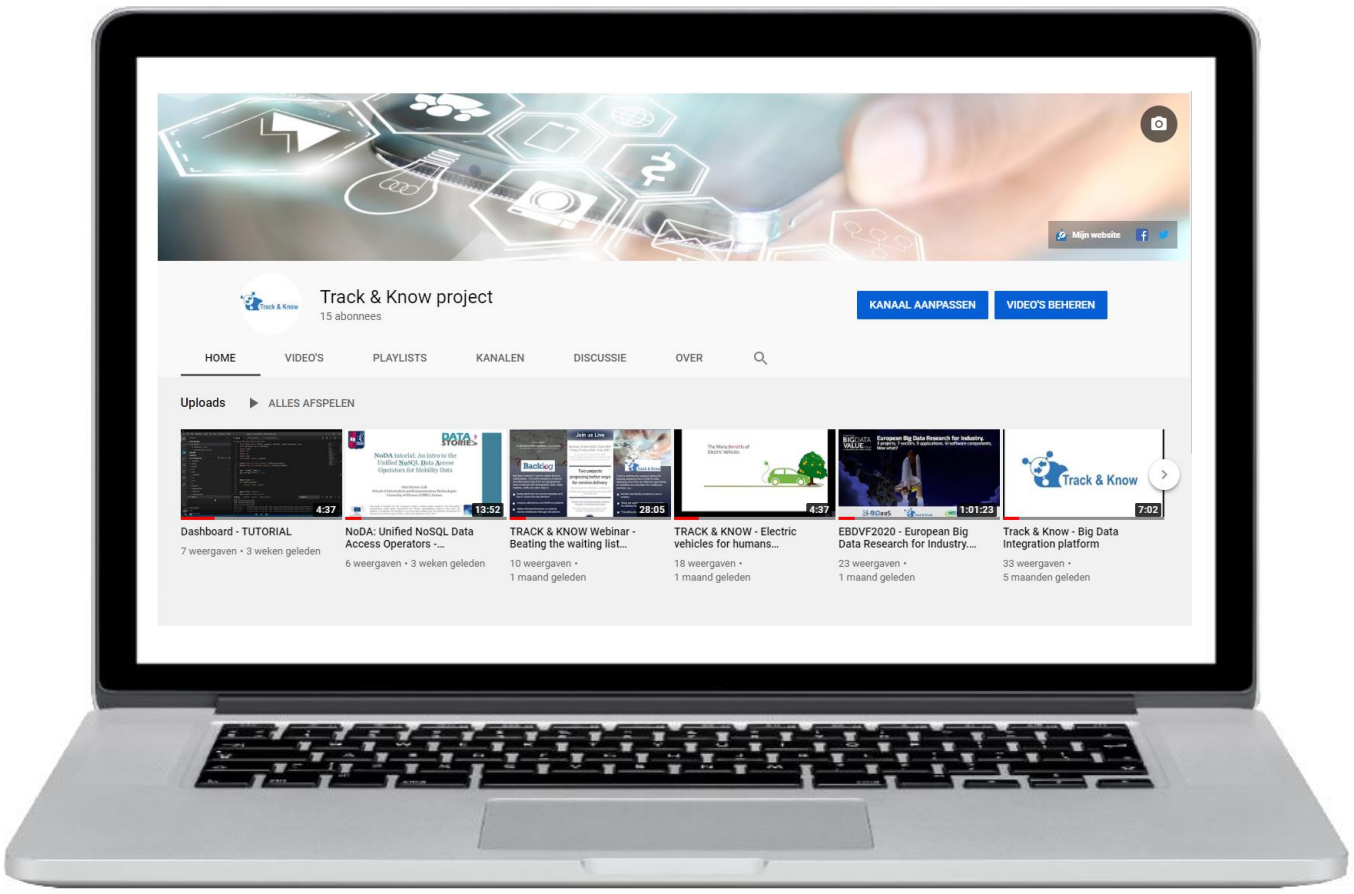
Are you in need of any additional info? Do not hesitate to browse through the **project website** and its **online observatory**, filled with Track & Know software, datasets and relevant literature.

[Discover the project website now]

[Dive right into the online observatory]

Track & Know also offers a lot of interesting **videos**. From the project video which gives an insight in the project's goals and challenges, to comprehensive, short explainer videos with an exclusive focus on a specific software tool. All videos can be found on the Track & Know YouTube channel, also including tutorials, webinars and video presentations.

[Go to the Track & Know YouTube channel]



RECENT PUBLICATIONS

- Adnan, M, Gazder U., Yasar A. H., Bellemans T., Kureshi, I (2020), Estimation of travel time distributions for urban roads using GPS trajectories of vehicles: a case of Athens, Greece, *Personal and Ubiqutious Computing*, DOI: https://doi.org/10.1007/s00779-020-01369-4, Download **here**
- Theodoridis Y. (2020) Learning from Our Movements – The Mobility Data Analytics Era. In: Tserpes K., Renso C., Matwin S. (eds) Multiple-Aspect Analysis of Semantic Trajectories. MASTER 2019. Lecture Notes in Computer Science, vol 11889. Springer, Cham. Download **here**
- Tampakis, P., Doulkeridis, C., Pelekis, N., Theodoridis, Y. (2020). "Distributed Subtrajectory Join on Massive Datasets." ACM Digital Library. Published version: https://dl.acm.org/doi/10.1145/3373642
- Nikitopoulos, P., Sfyris, G.A., Vlachou, A., Doulkeridis, C., Telelis, O. (2020) "Pruning Techniques for Parallel Processing of Reverse Top-k Queries." Distributed and Parallel Databases (Springer), 2020. Published version: DOI 10.1007/s10619-020-07297-9
- Guidotti, R., Nanni, M. (2020). Crash Prediction and Risk Assessment with Individual Mobility Networks. The 21st IEEE International Conference on Mobile Data Management (MDM 2020). **To appear soon.** Pre-print version **here**.
- Yeghikyan, G., Opolka, F., Lepri, B., Nanni, M., Lio, P. (2020) Learning Mobility Flows from Urban Features with Spatial Interaction Models and Neural Networks. IEEE International Conference on Smart Computing (SMARTCOMP). **To appear soon.** Pre-print version **here**.
- Liu, F., Andrienko, G., Andrienko, N., Chen, S., Janssens, D., Wets, G., Theodoridis, Y. 2020. "Citywide traffic analysis based on the combination of visual and analytic approaches." Download **here**
- Andrienko, N., Andrienko, G. "Spatio-temporal visual analytics: a vision for 2020s." Journal of Spatial Information Science, 2020, number 20, pp.87-95. Download **here**
- Tritsarolis A, Theodoropoulos GS, Theodoridis Y (2020). "Online discovery of co-movement patterns in mobility data." Int. J. Geographical Information Science, Taylor & Francis. **Pre-print available here**
- Andrienko, G., Andrienko, N., Patterson, F., Chen, S., Weibel, R., Huang, H., Doulkeridis, C., Georgiou, H., Pelekis, N., Theodoridis, Y., Nanni, M., Longhi, L., Koumparos, A., Yasar, A. and Kureshi, I. "Visual Analytics for Characterizing Mobility Aspects of Urban Context." Wenzhong Shi, Michael Goodchild, Michael Batty, Mei-Po Kwan, Anshu Zhang (Eds.) **Urban Informatics**. Springer, 2020. **Pre-print available here.**
- Andrienko, N., Andrienko, G. "Visual Analytics of Vessel Movement." Alexander Artikis and Dimitris Zissis (Eds.) **Maritime Informatics**. Springer, 2020. **Pre-print available here.**
- Andrienko, N., Andrienko, G., Fuchs, G., Slingsby, A., Turkay, C., Wrobel, S. **Visual Analytics for Data Scientists**. Springer, 2020. **Published version available here.**

Full list of publications 2018-2019-2020 is available HERE

Find & Follow us

Website: https:// trackandknowproject.eu

Twitter: @TrackandKnow

LinkedIn: Track & Know

The project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No 780754.